# Sequence, Structure and Evolution of Metazoan OVO-like Genes.

Abhishek Kumar[1,2,*], Anita Bhandari[3], Miss. Sushma[4],

Umesh Kumar[5] & Pankaj Goyal[6,7]

[1]Department of Genetics & Molecular Biology in Botany, Institute of Botany, Christian-Albrechts-University at Kiel, Kiel, Germany

[2]Dept of Biology, University of Padova. via U.Bassi 58b · I-35131 Padova Italy.

[3]Molecular Physiology, Zoological Institute, Christian-Albrechts-University at Kiel, Kiel, Germany

[4]Indian Institute of Toxicology Research, Lucknow, Uttar Pradesh, India

[5]Department of Biology, Carleton University, Ottawa, Ontario, Canada

[6]Institute for Prevention of Cardiovascular Diseases, University of Munich, D-80336 Munich, Germany

[7]Current Address: Department of Biotechnology, Central University of Rajasthan, Bandar Sindri, Rajasthan, India

*Corresponding author, AK: Abhishek.abhishekkumar@gmail,com

## Abstract

A hierarchy set of transcriptional regulators controls the development of multicellular organisms by sequential activation by transcription factors such as OVO-like proteins (belongs a zinc finger family). OVO-like genes of mammals are considered as orthologs of OVO gene from *Drosophila melanogaster* until now. Using comparative bioinformatic analysis of OVO-like genes of vertebrates, we have identified three different types of OVO-like genes OVOL1-OVOL3 in vertebrates and a paralog of OVOL3 - OVOL4 is detected in fish lineages. Upon comparing the domains of these OVO-like proteins from different metazoa genomes, we found that there is basal domain consisting tetrad of C2H2 zinc fingers to which by N-/ C-extensions various types of OVO-like genes/proteins are evolved in different lineages of metazoan where size of extensions varied from hundreds to several hundreds of amino acids and these extensions do not share homologies with OVO-like genes from placozoans to mammals. By corroborating the full length domains of OVO-like proteins, it is clear that human OVO-like proteins OVOL1-OVOL3 are merely homologs of *Drosophila* OVO, but not ortholog as until now described in databases.

# Introduction

Transcriptional regulatory activities are assisted by transcription factors such as a family of zinc finger proteins possessing zinc finger motifs in their core domain. OVO-like proteins function as transcription factors to regulate gene expression in various differentiation processes [1-3]. Drosophila ovo is a prototype of ovo genes, which is most extensively characterized [4-5], with multiple spliced isoforms encode at least four protein isoforms (A-D), all containing four identical Cys2/His2 (C2H2) zinc fingers at C-terminal end while isoforms differ at N-terminal end. A homolog of OVO-like gene, *lin-48* from *Caenorhabditis elegans* encodes a C2H2 zinc-finger protein similar to the product of the *Drosophila ovo* gene [2]. OVO-like proteins primarily act as either transcriptional activators or transcriptional repressors [3,6-9], a typical function of zinc finger motif carrying family of proteins. For the purpose functional studies, OVO like genes are primarily reported in human and in mouse such as OVOL1, OVOL2 and OVOL3. Various functional studies in selected model organisms (such *D. melanogaster, C. elegans* and mice) corroborated that OVO genes are involved in the development and differentiation of a number of epithelial lineages [2,6,10-16]. Ovol1 is auto-repressor of expression by counteracting c-Myb activation and histone acetylation of its own promoter [11]. Ovol2 is identified as a key regulator of neural development in mice [13]**.** To date, there is no functional study reported for mammalian OVOL3.

Moreover, very little is known about their molecular evolution of OVO like genes in vertebrate lineage until now. Herein, we unravel molecular evolutionary insights of vertebrate OVO like genes due to two main reasons. First, there is no information yet about functional roles of OVO like genes from non-mammalian vertebrates. Second, it is expected that protein families diverged in fishes for genetic and functional novelties due to whole genome duplication which occurred in fish lineage after its separation from fishes [17-18] and it is interesting quest to find out what genomic novelties are associated with OVO like proteins in fishes. Recently, the genomes of a variety of organisms have been completely or nearly completely sequenced, facilitating the identification of OVO like genes in different vertebrate species.

In present work, using searches based on the conservation of nucleotides, genomic fragments and amino acidic sequences and domain architecture, we have identified *in silico* the putative OVO like genes of the following vertebrates: five teleost species (*Tetraodon nigroviridis* – Tetraodon [19], *Takifugu rubripes* – Fugu [20], *Oryzias latipes* – medaka [21], *Gasterosteus aculeatus* -stickleback and *Danio rerio* - Zebrafish), one amphibian species (*Xenopus tropicalis* - western clawed frog [22]), three avian species (*Gallus gallus* – chicken [23], *Taeniopygia guttata* - zebra finch [24] and *Meleagris gallopavo* - turkey), one reptile species (*Anolis carolinensis* - Anole lizard) and four mammalian representatives (*Homo sapiens* – human [25], *Mus musculus* – mouse [26], *Rattus norvegicus* – rat [27] and *Monodelphis domestica* - Opossum). Furthermore, we extended our analysis in different metazoan genomes such as lancelets - *Branchiostoma floridae* [28], sea urchin - *Strongylocentrotus purpuratus* [29], flies –*Drosophila melanogaster* [30], worms - *Caenorhabditis elegans* [31], annelids - *Helobdella robusta,*

and molluscs - *Lottia gigantea*. The information/noise ratio in protein sequence alignments is better in compared to alignments of DNA because of the fact that the proteins are built from a repertoire of twenty variables - amino acids while DNA only contains four different bases [32-33]. Putative homologs of vertebrate OVOL genes from public genome database were recognized by homology searching tools such BLAST suite [34-36] or FASTA suite [37-38]. Orthologies were confirmed by in silico comparative genomic methods, such as conserved gene synteny between different vertebrate species teleosts vs. amphibians vs. birds vs. mammals, phylogenetic analysis with orthologues from different species, and functional mapping of orthologues, using a set of amino acid residues known to serve critical roles in functions of the proteins. Comparative analyses using different vertebrate species were used to study the evolution of OVOLs.

# Results

## *Orthology assessment of OVOL1 in vertebrates*

Conservation of genomic organization of a gene on a given chromosome or scaffold provides a useful tool for predicting functional interaction. Selective processes are essential to preserve the organization of these clusters in closely related organisms. Thus, chromosomal localization and gene order conservation is a vital tool in assigning orthology for a given set of genes in a gene family. We compared the syntenic organization of OVOL1 genes from different vertebrates. OVOL1 gene is localized on chromosome 11 in human genome as shown in **Figure 1**. There is set of genes flanking OVOL1 gene in both side such as triad of SIPA1-RELA-KAT5 on the one side and a set of five genes namely SNX32-MUS81-RIBP-FOSL1-BANF1 on the other side in a region of ~380 kb. This syntenic architecture is maintained across several mammals such as in mouse (chromosome 19/400 kb fragment), in rat (chromosome 1/300 kb fragment), and in opossum (chromosome 8/300 kb fragment). However, when this fragment is searched into bird genomes such as chicken, turkey, and zebra finch, we do not find OVOL1 genes at all. Additionally, complete genomic locus comparable to mammalian genome fragment is also missing in these genomes (**Figure 1**).

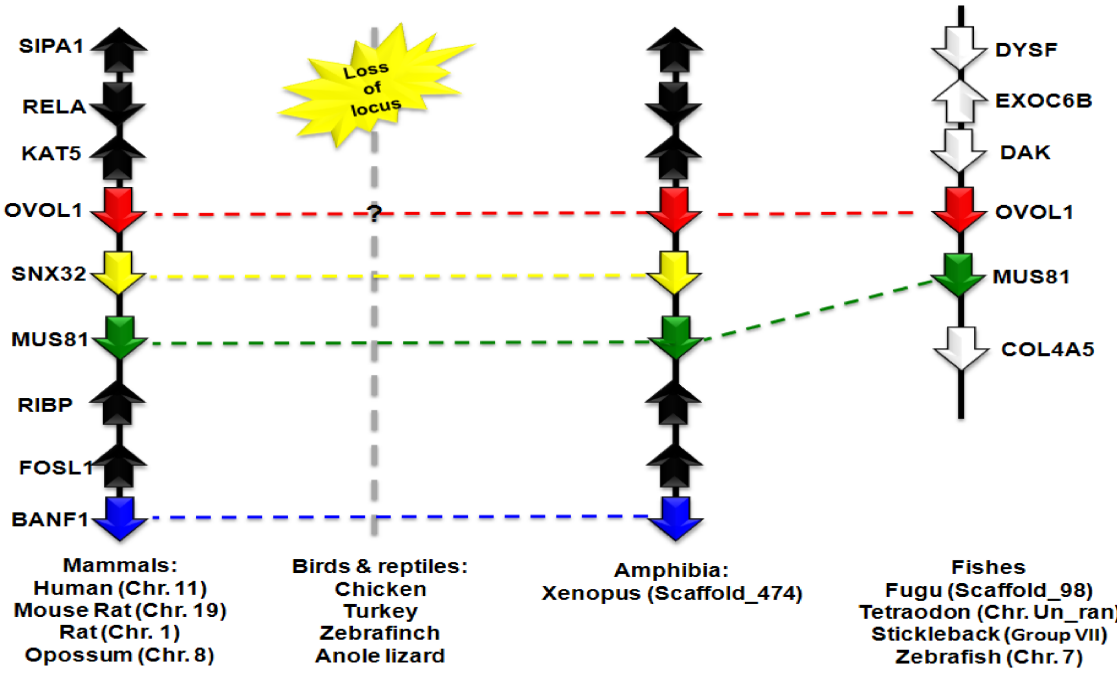


**Figure 1. Syntenic analysis of OVOL1 gene from selected vertebrates, flanked by a set of conserved marker genes.**

Subsequently, we found same results as of birds in only sequenced reptile genome – Anole lizard, *A. carolinensis* (AcoCar1.0 assembly). However, we traced OVOL1 gene in the frog *X. tropicalis* genome with identical locus as of mammals on scaffold_474 in a 900 kb fragment. We identified OVOL1 genes in different fish genomes flanking a triad of DYSF-ECOC6B-DAK on one side and MUS81-COL4A5 on the other side in a fragment of 300 kb, 340 kb, 350kb and 380 kb and in *Fugu* (scaffold_98), *Tetraodon* (chromosome Un_random), stickleback (groupVII) and zebrafish (chromosome 7), respectively. Although the sets of marker genes are largely varied in tetrapod and fishes, however, presence of a single copy of highly conserved gene - MUS81 (encodes for 611 amino acid long Crossover junction endonuclease MUS81) in all genomes of vertebrates, concludes that these OVOL1 locus is orthologically conserved from fish to human. OVOL1 orthologs from different vertebrates are listed in **Table 1**.

## *Orthology assignment of OVOL2 in vertebrates*

Upon tracing OVOL2 orthologs in different vertebrate genomes, we found that OVOL2 gene is localized on chromosome 20 in the human genome flanking a set of marker genes such as a triad of RRBP1-BANF2-SNX5 on one side and a set of five genes – RP2BP-POLR3F-RBBP5-SEC23B-DTD1 on the other side in a region of ~900 kb (**Figure 2**).

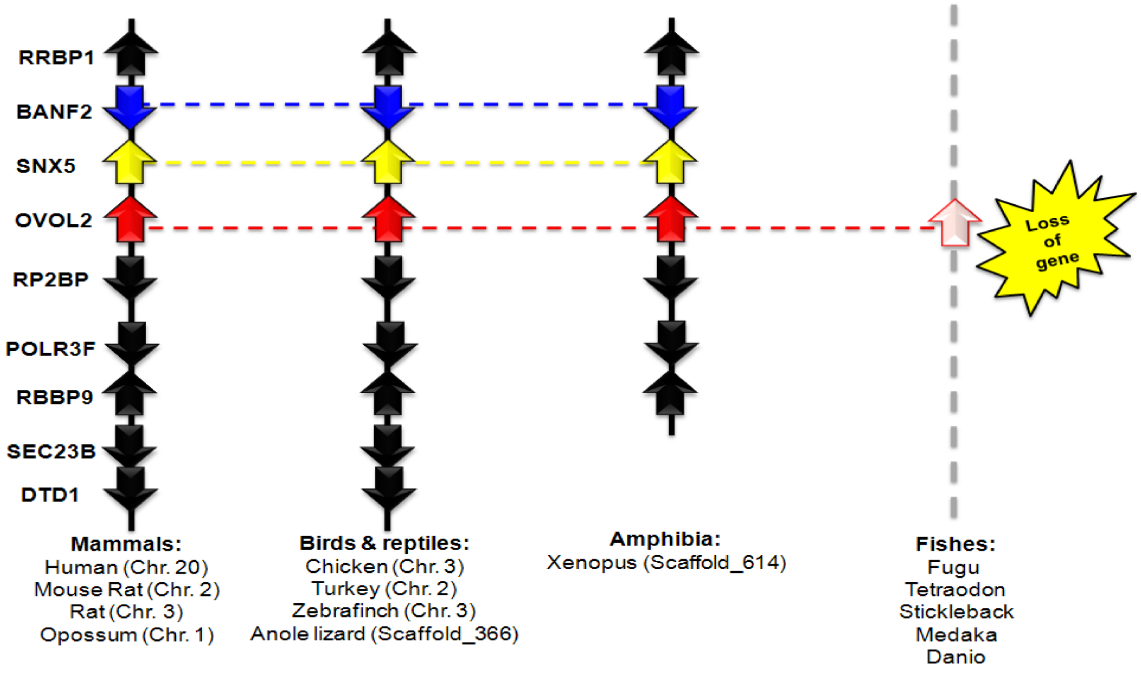


**Figure 2. Vertebrate OVOL2 orthologs identified by comparing chromosomal localization of this gene from evolutionary important vertebrates.**

This genomic fragment is maintained in large sets of mammals such as in mouse (chromosome 2/~600 kb), in rat (chromosome 3/600 kb fragment) and in opossum (chromosome 1/1.3 Mb fragment). By comparing this genomic architecture in bird genomes, we detected a similar fragment as reported above for mammalian OVOL2 gene for OVOL2 gene from birds with fragment size of 220 kb, 200 kb and 200 kb in chicken (chromosome 2), zebra finch (chromosome 3), and turkey (chromosome 2) , respectively.

Furthermore, we have identified this syntenic organization in anole lizard on scaffold_366 in a region of about 200 kb. However, we could not detect OVOL2 gene in fish genomes although we could find marker genes scattered on different locus. This corroborates that OVOL2 gene is missing in five analyzed fish genomes. OVOL2 orthologs from different vertebrates are listed in **Table 2**.

## *Unraveling human OVOL3 orthologs in different vertebrates*

While tracing the OVO like genes, we identified third gene OVOL3 in wide array of mammals such as human (chromosome 19), chimpanzee (chromosome 19), mouse (chromosome 7), rat (chromosome 1), cow (chromosome 18), pig (chromosome 6), and opossum (chromosome 4) with a conserved synteny. In this conserved synteny, OVOL3 gene is flanked by an octet of LIN37-PRODH2-KIRREL2-APLP11-NKF3ID-LPFN3-SDHAF1-CLIF3 on one side and a triad of POLR2L-CAPSN1-COX7A1 on the other side in a region of ~400 kb in human chromosome 19 (**Figure 3**).

In fishes, this genomic arrangement is not found, however, there is another genomic organization on which OVO like gene is localized, which we named as OVOL4. OVOL4 is flanked by a tetrad of AMOT-HLCS-REXO2-DMPK and by AKT2-2 on the other side on scaffold_455 in *Fugu.* Similar architecture is maintained in zebrafish (chromosome 10) and in medaka (chromosome 14). On reverse searching fish marker genes in mammalian genomes, we found that AKT2 gene that encode for an enzyme RAC-beta serine/threonine-protein kinase and this gene has two copies in fishes as AKT2-1 and AKT2-2. We detected AKT2-1 is found on close to mammalian cluster of OVOL3, suggesting that OVOL4 fishes is paralogous to OVOL3 of mammals. OVOL3 orthologs

from different vertebrates are listed in **Table 3** and OVOL4 genes of fishes are listed in **table 4**.

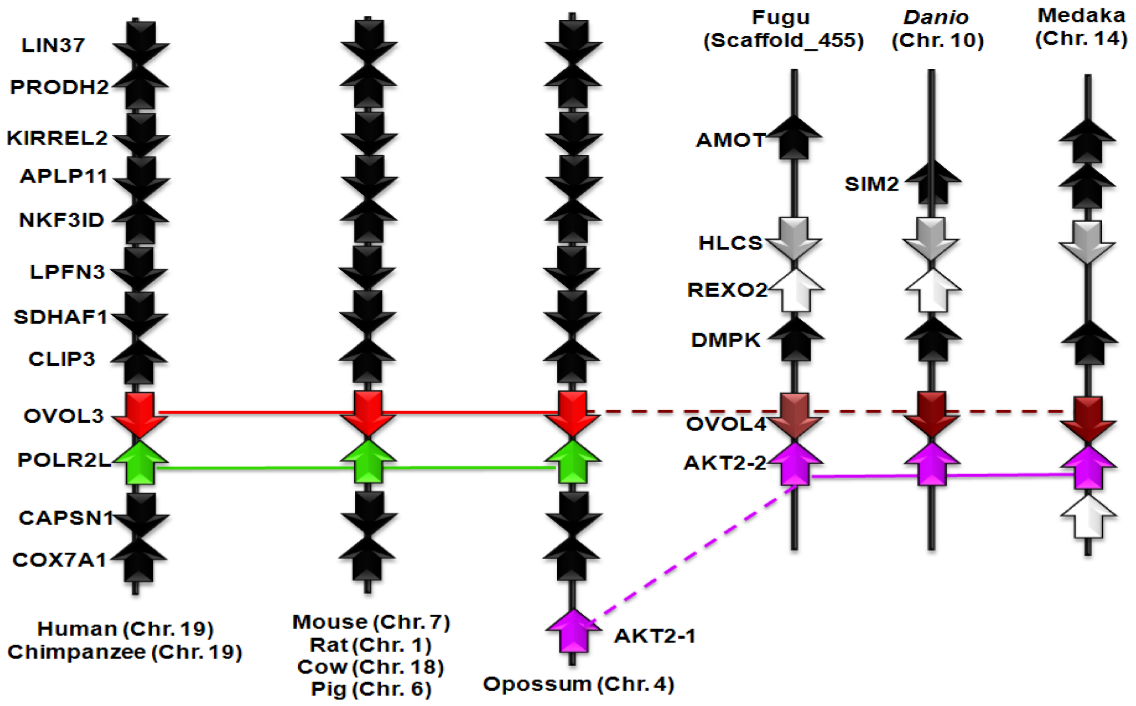


**Figure 3. Syntenic conservation of OVOL3 gene and its paralog OVOL4 from selected vertebrates.**

On closely inspecting chromosomal localizations of OVOL1-OVOL3 from vertebrates, we found that syntenies of OVOL1 and OVOL2 share marker genes that homologous marker genes such as Barrier-to-autointegration factor encoding genes - BANF1 and BANF2 (**marked in blue in Figure 1 and Figure 2**), respectively and sorting nexin homologs - SNX32 and SNX5 (**marked in yellow in Figure 1 and Figure 2**),, respectively. This indicates that OVOL1 and OVOL2 are originated by fragmental duplications before 450 MY ago since it is maintained from fish to mammals. Surprisingly, birds have only one copy of OVO like genes – OVOL2, indicating either there is bird specific adaptations do not require second copy of OVO like genes or OVO

like genes have similar physiological roles which can be performed by any of three OVO like genes.

## Evolutionary history of OVO like domains in metazoan

Upon protein sequence analysis of OVO like (OVOL) proteins from vertebrates, we found that OVOL1, OVOL2, OVOL3 and OVOL4 share a common domain with highly conserved tetrad of C2H2 type zinc finger motif in C-terminal region. The peptide length of OVO like (OVOL) proteins varied from 215-286 amino acids. OVOL1 has larger peptide length and OVOL3 and OVOL4 has smaller peptide length due to variability in N- and C- terminal extension. Secondary structure of human OVOL1 is predicted using PSIPRED [39] as depicted in **Figure 4**.

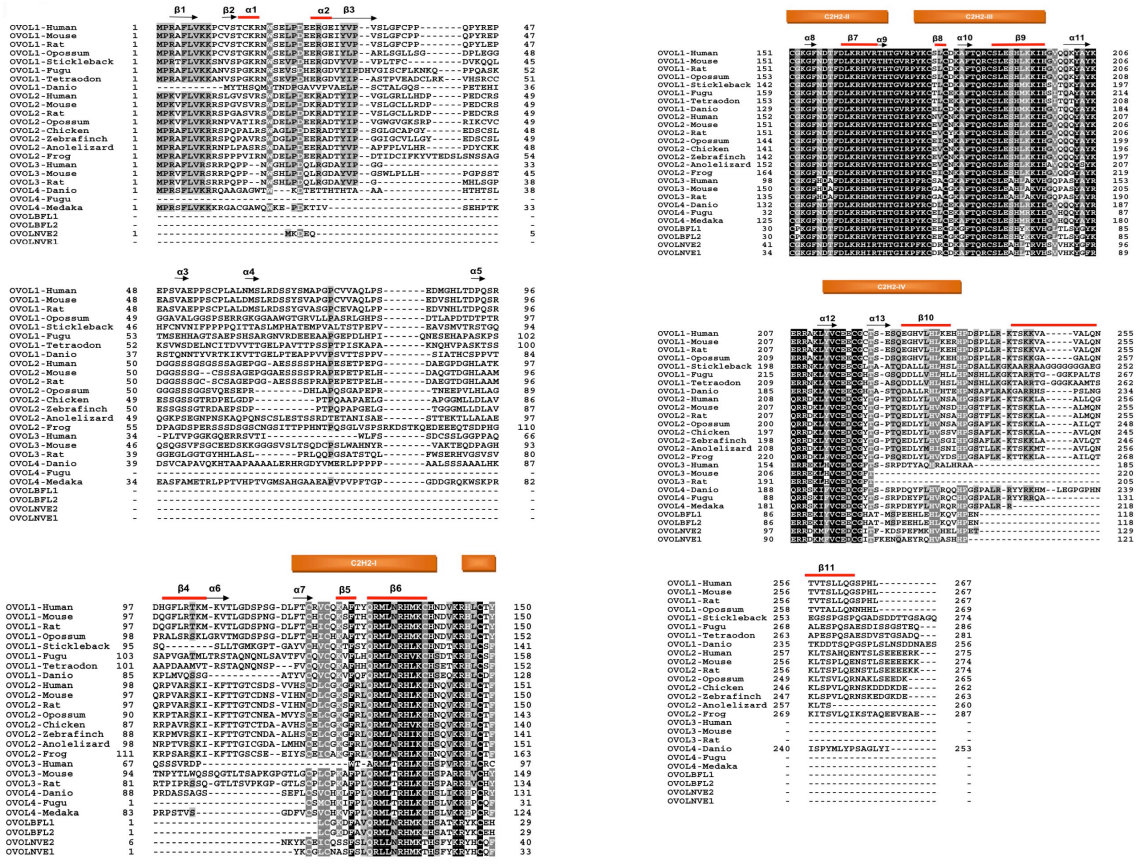

**Figure 4. Alignment of OVO like proteins from different vertebrates, *B. floridae and N. vectensis.*** This alignment is created by MUSCLE [64-65] and further edited in GeneDoc [66]. Secondary structures of human OVOL1 is predicted using PSIPRED [39], and are marked above the alignment. Four C2H2 zinc finger motifs (I-IV) are marked by orange bar. The rodent OVOL3 protein terminates at position 10 in C2H2 motif IV.

There are thirteen α-helices and eleven β-sheets are present in human OVOL1 protein. There are small stretches of disordered regions in first 100 amino acids from N-terminal regions of OVOL1-OVOL3 proteins (**Figure 5A-5C**) as predicted by DISOPRED2 software [40].

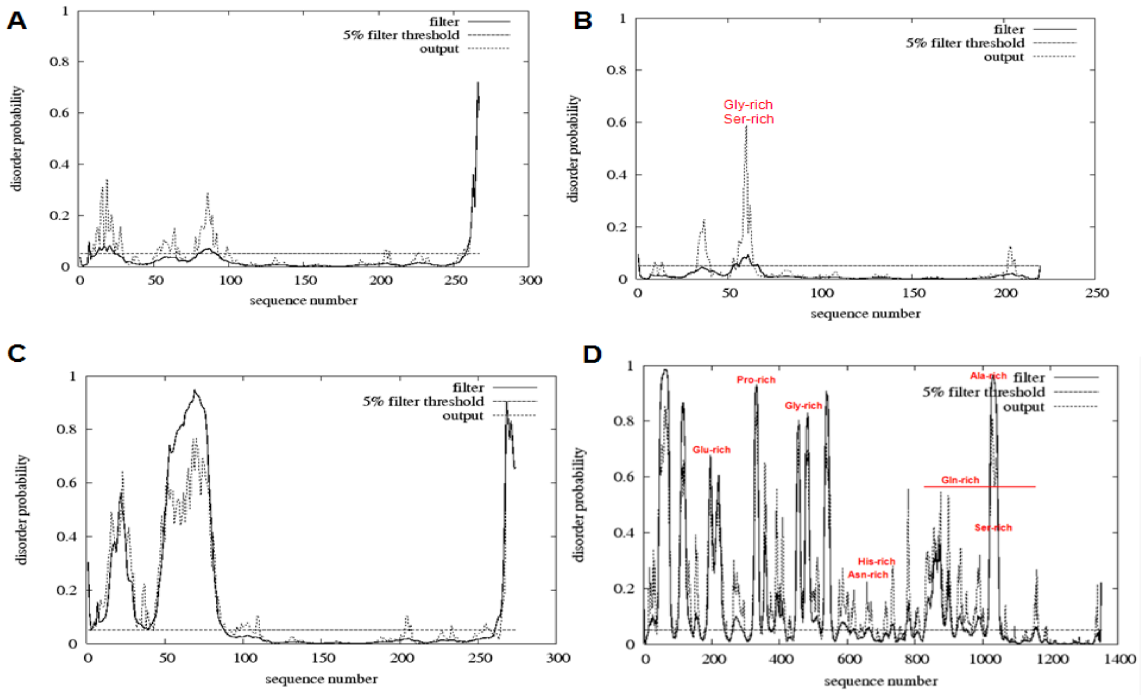


**Figure 5. Presence of disordered regions in different OVO like proteins.**

**A.** Mouse OVOL1 has disordered residues in first 100 amino acids.

**B.** Mouse OVOL2 possesses disordered residues in first 50 amino acids with a glycine-rich and Serine rich region as marked in red color.

**C.** Mouse OVOL3 has disordered segments in N-terminal 100 residues.

**D.** *Drosophila* OVO is intrinsically disordered with large patches of residue biasness as indicated by red color.

This prediction is obtained by DISOPRED2 software [40]. The horizontal line is the order/disorder threshold for the default false positive rate of 5%. The 'filter' curve represents the outputs from DISOPRED2 and the 'output' curve the outputs from a linear SVM classifier (DISOPREDsvm). The outputs from DISOPREDsvm are included to indicate shorter as low confidence predictions of disorder.

There are four C2H2 type zinc finger motifs in mammalian OVO-like proteins in

following regions 118 – 140 (23 aa), 146-168 (23 aa), 174-197 (24 aa), and 213-236 (24), numbering according to human OVOL1. This domain is highly conserved in different OVO like proteins from wide array of vertebrates. Normally *Drosophila* OVO is considered ortholog of human or mouse OVO like genes, However, there is a word of caution in such cases as *Drosophila* OVO gene encodes for four alternatively spliced isoforms names as A-D. *Drosophila* OVO-B isoform encodes for full length peptide length of 1351 amino acids, whereas size of isoforms A, C and D are 975, 1222 and 1028 amino acids, respectively. *Drosophila* OVO has four C2H2 zinc finger motifs at following positions - 1197 – 1219 (23 aa), 1225 – 1247 (23 aa), 1253 – 1276 (24 aa), and 1292 – 1315 (24 aa), respectively at C-terminal ends. Furthermore, The full-length *Drosophila* OVO protein is further characterized by presence of amino acid biasness or low complexity regions which is called as disordered regions as predicted by DISOPRED2 software [40]. *Drosophila* OVO protein appeared to be intrinsically disordered (**Figure 5D**) with following biasness for specific amino acids - a Glu-rich region between 196 – 239 (44 aa), a Pro-rich between 309 – 342 (34 aa), a Gly-rich between 448 – 618 (171 aa), an Asn-rich between 620 – 660 (41 aa), a His-rich region between 645 – 65 (39 aa) , Gln-rich region between 837 – 1158 (322 aa), and an Ala-rich region between 1001 – 1059 (59 aa) and a Ser-rich region between 1025 – 1045 (21 aa), respectively. Vertebrate OVO-like proteins and *Drosophila OVO* share strong conservation in C-terminal end possessing tetrad of C2H2 zinc finger motifs.

To further delineating OVO like proteins from different metazoan, we have traced OVO like genes in a set of metazoa genomes using BLAST suite [34-36]. Based on homology searches for human OVO like genes, we identified two genes from *Branchiostoma*

*floridae* genome sharing high degree of conservation with JGI accession id e_gw.374.48.1 and e_gw.236.92.1. Furthermore, on searching these genes in sea urchin - *Nematostella vectensis* genome, we identified two genes that possess higher conservation with human OVO domains with JGI accession id gw.31.97.1 and e_gw.31.122.1, respectively. OVO like proteins these two species are comprised of a single domain with tetrad of C2H2 zinc finger motif as possessed by vertebrate OVO like proteins (**Figure 6**).

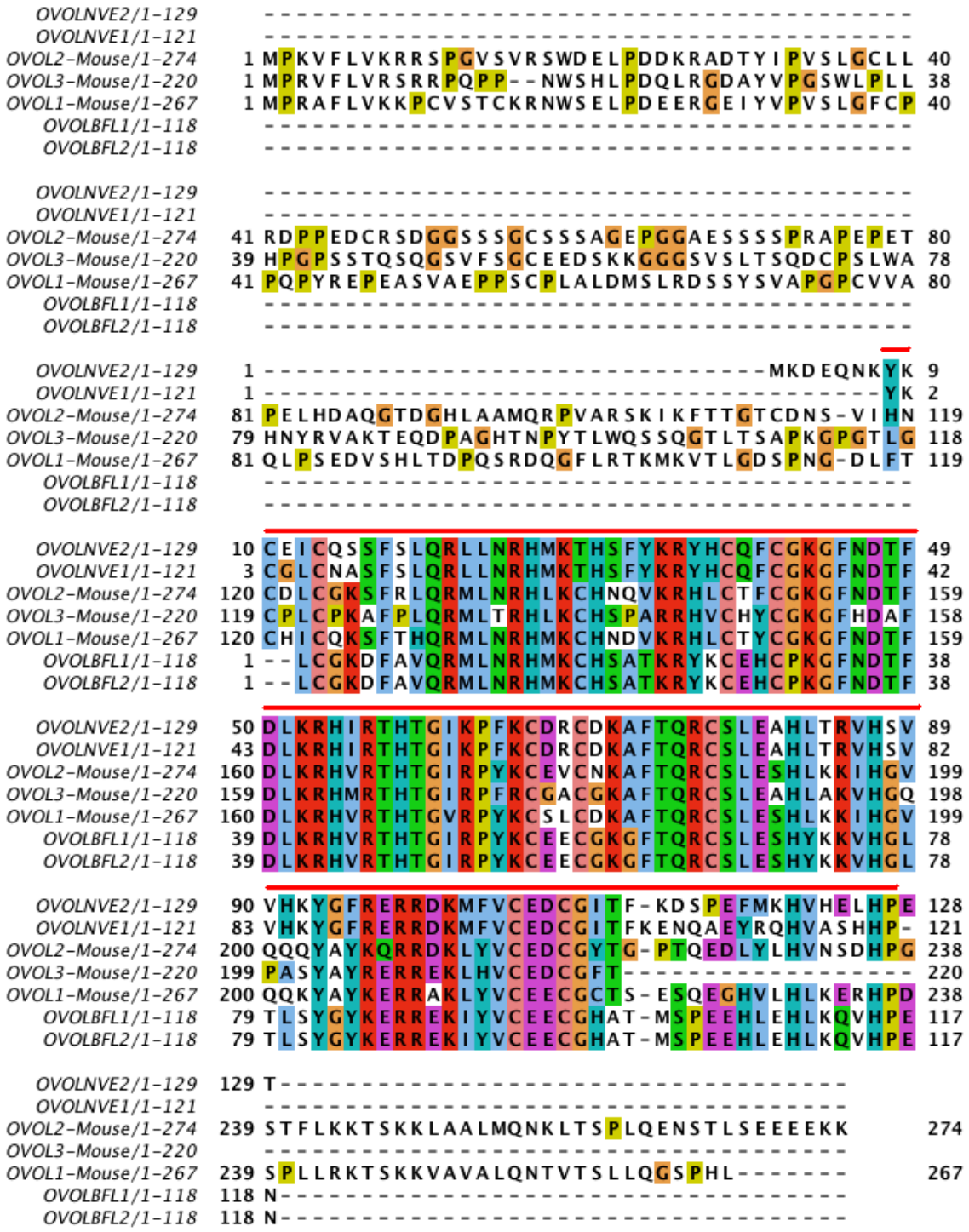


**Figure 6. Protein Alignment of OVO like proteins from mouse, lancelets and sea anemone depicting conservation in C-terminal end possessing tetrad of C2H2 zinc finger motifs (marked by red bar).** This alignment is created by MUSCLE [64-65] and further edited by Jalview [67-68].

To understand the molecular evolution of OVOL like proteins, a phylogenetic tree

(**Figure 7**) of OVO like proteins was constructed using Neighbor-Joining method [41] with bootstrap value =1000 replicates [42] with help of MEGA4 software [43]. This phylogenetic tree demarked the different types of OVO like orthologs such as OVOL1, OVOL2 and OVOL3-OVOL4 along with basal *B. floridae* and *N. vectensis* OVO like proteins named as OVOBFL1/OVOBFL2 and OVONVE1/OVONVE2, respectively.

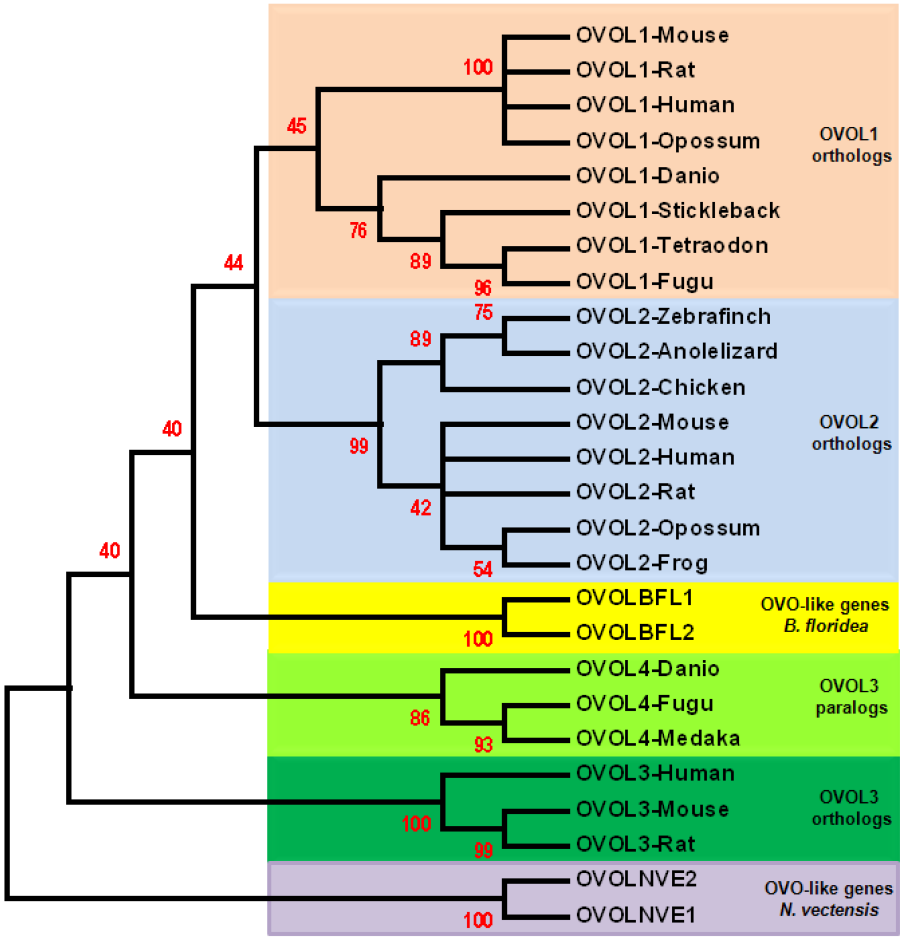


**Figure 7. Evolutionary history of vertebrate OVO like proteins using Neighbor-Joining method [41] with bootstrap value =1000 replicates.** Branches corresponding to partitions reproduced in less than 35% bootstrap replicates are collapsed. The percentage of replicate trees in which the OVO like proteins clustered together in the bootstrap test (1000 replicates) are shown next to the branches [42]. Phylogenetic analyses were conducted in MEGA4 [43]. Different colors indicates various types of OVO like genes such as OVOL1-OVOL3 orthologs and paralog of OVOL3 in fishes named as OVOL4. Furthermore, OVO like genes from lancelets and sea anemone are also clustered in the tree separately.

On further searching human OVO like proteins in the leech *H. robusta* genome, we identified, OVO like proteins in this annelid genome with accession id e_gw1.1.1891.1 and e_gw1.4.1162.1, respectively. In JGI Genome *L. gigantea* v1.0, fgenesh2_pg.C_sca_13000299 & e_gw1.13.34.1 are found as OVO like genes upon homology searches. Two OVO like proteins were detected from sea urchin – *S. purpuratus* genome with accession id XP_796652.2 and XP_788176.1. These two proteins from *S. purpuratus* share conserved C2H2 zinc finger carrying domain, in addition to non-homologous N-terminal extensions. On comparing four C2H2 type zinc finger motifs from sea anemone to vertebrates**,** we found that these four motifs are conserved **(Figure 8)**, and these are typical C2H2 type zinc finger motifs as reported description of these motifs in Prosite pattern database with accession number PS00028.

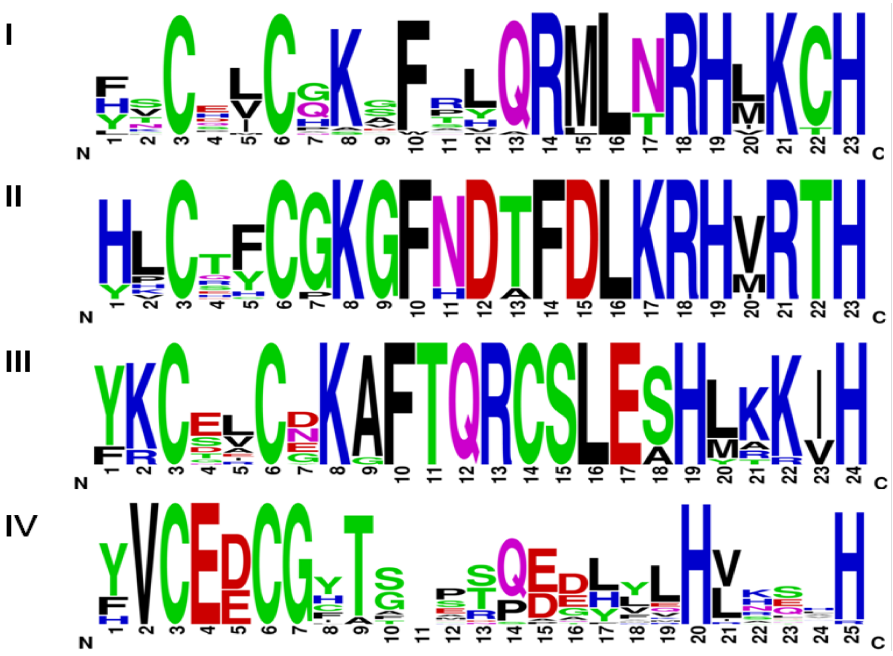


**Figure 8. Sequence logo of four different Cys2-His2 (C2H2) zinc finger motifs (I-IV) present in different OVO like proteins from sea anemone to human.** This sequence logo is generated from comprehensive protein alignment of OVO-like proteins (**supplementary figure S1**) using WebLogo 3.0 [69]. C2H2 zinc finger motif IV has 25 amino acids due to presence of one extra amino acid OVOLNVE1 protein from sea anemone, which is at eleventh position in sequence logo of C2H2 zinc finger motif.

**Figure 9** depicts domain evolution of OVO like proteins from different lineage of metazoan over period of > 700 million years.

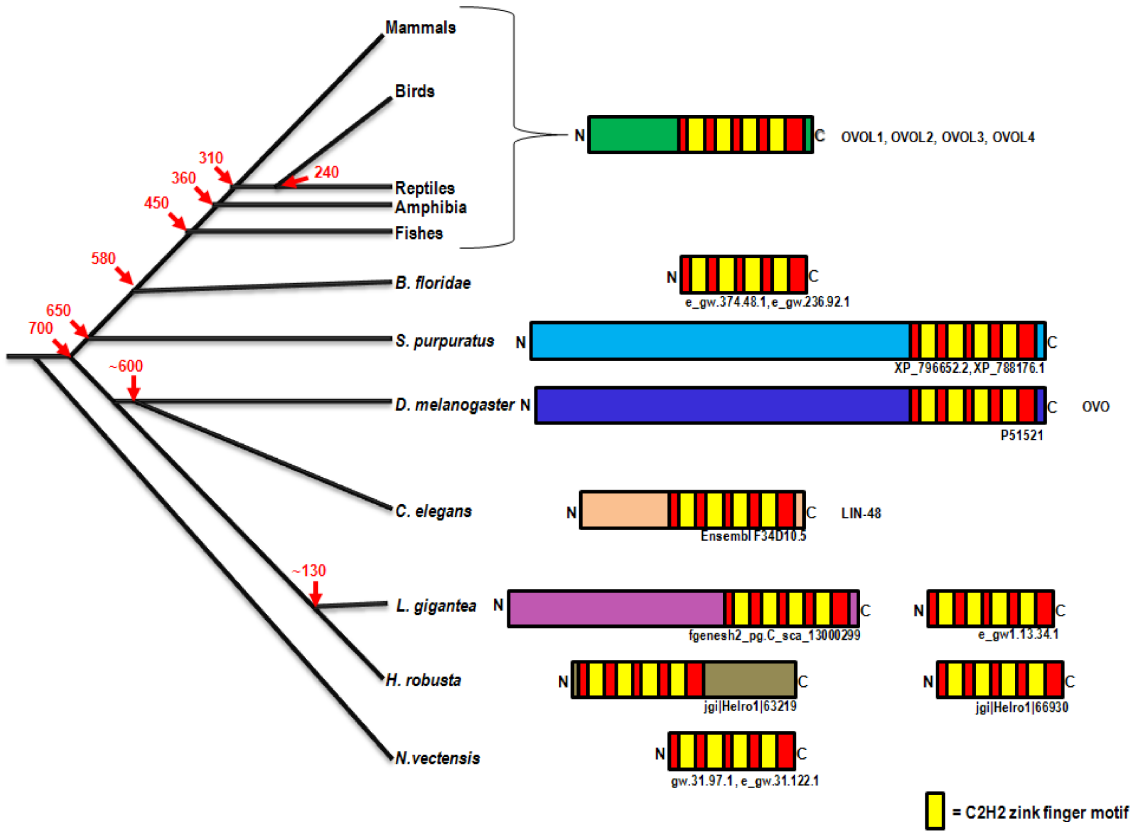


**Figure 9. Protein domain evolution of OVO like proteins from different lineage of metazoan over period of > 700 million years.** A highly conserved domain of tetrad C2H2 zinc finger motifs (red and yellow box) is found in all evolutionary important organisms. Primarily N-terminal extensions in C2H2 lead to different types of protein products with exception in jgi|Helro1|63219 in leech *H. robusta* where extension is found in C-terminal end of C2H2 zinc finger motif. The time period is marked with help of works of Kumar and Hedge (2003) [70] and Ponting (2008) [71].

The basal metazoan such sea anemone possesses only C2H2 zinc finger carrying OVO like domain, to which by N-/C-terminal extensions lead into different types of OVO like proteins, predominantly these extensions are N-terminal with exception of jgi|Helro1|63219 in leech *H. robusta* where extension is found in C-terminal end of C2H2 zinc finger motif. The extension peptides varied from hundred to several hundreds of amino acids such vertebrate OVO like proteins and LIN48 from *C. elegans* have 100-120

amino acid extension with disordered region (**Figure 10**). Whereas *Drosophila* OVO and *S. purpuratus OVO* like proteins has several hundreds of amino acid extension in N-terminal end. The extended amino acid regions do not share homology with other OVO like proteins from evolutionary distant organism. This corroborates that these proteins from different lineages are actually only "homologs" not ortholog of *Drosophila* OVO proteins as described in annotations of different databases at the moment, since it is expected that full length domains of orthologous proteins are conserved.

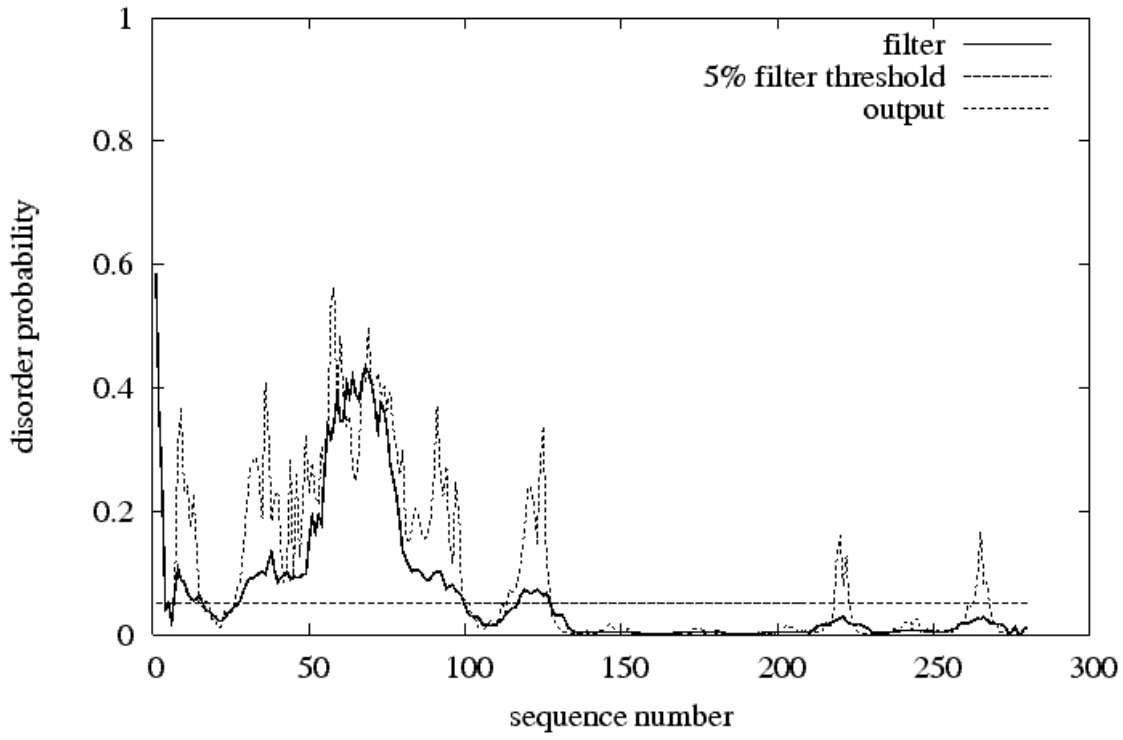


**Figure 10. LIN48 protein from *C. elegans* is characterized by presence of disordered segment of amino acids in first 100 residues as predicted by DISOPRED2 software [40].** Further details in Figure 4.

# Discussion

This study provides for the first time a comprehensive description of gene sequences, structural inputs, and detailed molecular phylogenetic studies of vertebrate OVO like genes. We have identified orthologs human OVO like genes from fish to mammals over 450 MY. Birds have only one copy of OVO like gene – OVO2 where as fishes have two genes OVOL1 and paralog of OVOL3 – named here as OVOL4. Presence of homologous marker genes in close vicinity of OVOL1 and OVOL2 loci correlates that OVOL1 and OVOL2 are originated by fragmental duplications. These fragments are maintained from fish to human, thus these are duplicated before separation of fishes from tetrapod lineage about 450 MY ago. The differential presence of OVO like genes in birds, reptiles and fish lineages, raises points to their role in development and differentiation of a number of epithelial cells in animal specific requirements such birds needed to reduce their body mass, thus do not need as many as mammalian genes. A regulatory mechanism is applied to development of epithelial cells and tissues as different organisms have adopted a different role such as flying and/or swimming under water.

Prior to this work, *Drosophila* OVO was considered to be ortholog of mammalian OVO like genes in different databases such as Ensembl V58 [44-45], but due to difference peptide compositions and peptide length it is not possible to assign orthology of *Drosophila OVO* in different metazoan lineages. However, OVO like proteins from different metazoan lineages, possess a conserved domain consists of four C2H2 zinc finger motifs to which by predominantly by N-terminal extensions with disordered segments, complete domain of varied OVO like proteins are originated over period of 700 MY. Only one case of C-terminal extension is found in *H. robusta.* These extended

regions in different OVO like proteins do not share significant homology. This extension of OVO like proteins are primarily consists of disordered residues, constituting a non-foldable domain with patches of multiple occurrence of same residue. In post-genomic era, there are several eukaryotic genomic sequences are available and from encoded proteins from these eukaryotes, it is evident that these disordered proteins are surprisingly common and these are frequently found in different eukaryotes. Furthermore with advent of genomic sequencing technologies and experimental methods, involvement of these domains become clear that these domains are found in many functional proteins [46-49] such as regulatory processes - cell signaling proteins [50-52] and transcriptional regulators [53].

These disordered segments have variable size from few amino acid sequences to entire domain (of several hundred amino acids), to even the entire protein as big as ~200 kDa [54]. Since, OVO like proteins are also transcriptional regulators, such disordered segments of varied length are also feature of such a regulatory system. Several of known disordered proteins or segments significantly differ in terms of sequences homology [54]. Thus, extended segments of OVO like proteins fall into same club of non-homologous disordered protein segments or domains.

This study enhances present understanding of OVOL genes from metazoan genomes. Furthermore, this study provides a good platform for those who are interested in characterizing OVO like genes from diverged species and also in vertebrate model systems such as *Xenopus, Gallus* and *Danio.*

# Materials and Methods

## *Data sources of genomic, cDNA and protein sequences*

The genomic DNA/cDNA/protein sequences from different eukaryotes were extracted via BLAST suite [55] searches using human/mouse OVOL1 as query sequence from the different genome databases such as human, mouse, rat and zebrafish genomes from National Centre for Biotechnology Information (NCBI) [56], *Takifugu , Xenopus tropicalis* and *Branchiostoma floridae* v1.0*, Helobdella robusta* v1.0, *Nematostella vectensis* v1.0 and *Lottia gigantea v1.0* from the DOE Joint Genome Institute (JGI) [57], *Tetraodon nigroviridis* from Ensembl [44-45] and the French National Sequencing Center (Genoscope) [58], and *Strongylocentrotus purpuratus* genome at the Human Genome Sequencing Center (HGSC), Baylor College of Medicine [59].

## Micro-synteny analysis across different genomes

To verify the orthology, micro-synteny across different genomes were analyzed using NCBI mapviewer [60], ENSEMBL genome browser [44-45], JGI genome browser [57], *Tetraodon* genome browser at the Genoscope [58] and UCSC genome browser [61].

## Sequence alignment of different OVO like proteins

Protein alignments of different OVO like proteins were generated with CLUSTALX 1.83 [62-63] or MUSCLE [64-65]. The alignments were edited and visualized different sequence characteristics using GENEDOC [66] or JALVIEW [67-68].

## Acknowledgement

We thank Chitra Rajakuberan for editing this manuscript. PG thanks Bayerische Staatsbibliothek München for its support to carry our this work.

# References

[1] Li B., Mackay D. R., Dai Q., Li T. W., Nair M., Fallahi M., Schonbaum C. P., Fantes J., Mahowald A. P., Waterman M. L., Fuchs E., Dai X., The LEF1/beta -catenin complex activates movo1, a mouse homolog of Drosophila ovo required for epidermal appendage differentiation, *Proc Natl Acad Sci U S A*, 2002, 99, 6064-6069

[2] Johnson A. D., Fitzsimmons D., Hagman J., Chamberlin H. M., EGL-38 Pax regulates the ovo-related gene lin-48 during Caenorhabditis elegans organ development, *Development*, 2001, 128, 2857-2865

[3] Andrews J., Garcia-Estefania D., Delon I., Lu J., Mevel-Ninio M., Spierer A., Payre F., Pauli D., Oliver B., OVO transcription factors function antagonistically in the Drosophila female germline, *Development*, 2000, 127, 881-892

[4] Garfinkel M. D., Wang J., Liang Y., Mahowald A. P., Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity, *Mol Cell Biol*, 1994, 14, 6809-6818

[5] Mevel-Ninio M., Terracol R., Salles C., Vincent A., Payre F., ovo, a Drosophila gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning, *Mech Dev*, 1995, 49, 83-95

[6] Payre F., Vincent A., Carreno S., ovo/svb integrates Wingless and DER pathways to control epidermis differentiation, *Nature*, 1999, 400, 271-275

[7] Andrews J., Levenson I., Oliver B., New AUG initiation codons in a long 5' UTR create four dominant negative alleles of the Drosophila C2H2 zinc-finger gene ovo, *Dev Genes Evol*, 1998, 207, 482-487

[8] Mevel-Ninio M., Fouilloux E., Guenal I., Vincent A., The three dominant female-sterile mutations of the Drosophila ovo gene are point mutations that create new translation-initiator AUG codons, *Development*, 1996, 122, 4131-4138

[9] Li B., Dai Q., Li L., Nair M., Mackay D. R., Dai X., Ovol2, a mammalian homolog of Drosophila ovo: gene structure, chromosomal mapping, and aberrant expression in blind-sterile mice, *Genomics*, 2002, 80, 319-325

[10] Teng A., Nair M., Wells J., Segre J. A., Dai X., Strain-dependent perinatal lethality of Ovol1-deficient mice and identification of Ovol2 as a downstream target of Ovol1 in skin epidermis, *Biochim Biophys Acta*, 2007, 1772, 89-95

[11] Nair M., Bilanchone V., Ortt K., Sinha S., Dai X., Ovol1 represses its own transcription by competing with transcription activator c-Myb and by recruiting histone deacetylase activity, *Nucleic Acids Res*, 2007, 35, 1687-1697

[12] Nair M., Teng A., Bilanchone V., Agrawal A., Li B., Dai X., Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription, *J Cell Biol*, 2006, 173, 253-264

[13] Mackay D. R., Hu M., Li B., Rheaume C., Dai X., The mouse Ovol2 gene is required for cranial neural tube development, *Dev Biol*, 2006, 291, 38-52

[14] Li B., Nair M., Mackay D. R., Bilanchone V., Hu M., Fallahi M., Song H., Dai Q., Cohen P. E., Dai X., Ovol1 regulates meiotic pachytene progression during spermatogenesis by repressing Id2 expression, *Development*, 2005, 132, 1463-1473

[15] Oliver B., Perrimon N., Mahowald A. P., The ovo locus is required for sex-specific germ line maintenance in Drosophila, *Genes Dev*, 1987, 1, 913-923
[16] Mevel-Ninio M., Terracol R., Kafatos F. C., The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development, *EMBO J*, 1991, 10, 2259-2266
[17] Ohno S., Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999, *Semin Cell Dev Biol*, 1999, 10, 517-522
[18] Ohno S., *Evolution by Gene Duplication*. New York, NY: Springer Verlag. , 1970
[19] Jaillon O., Aury J. M., Brunet F., Petit J. L., Stange-Thomann N., Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A., Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B., Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S., Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B., Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V., Cruaud C., Duprat S., Brottier P., Coutanceau J. P., Gouzy J., Parra G., Lardier G., Chapple C., McKernan K. J., McEwan P., Bosak S., Kellis M., Volff J. N., Guigo R., Zody M. C., Mesirov J., Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M., Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C., Wincker P., Lander E. S., Weissenbach J., Roest Crollius H., Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype, *Nature*, 2004, 431, 946-957
[20] Aparicio S., Chapman J., Stupka E., Putnam N., Chia J. M., Dehal P., Christoffels A., Rash S., Hoon S., Smit A., Gelpke M. D., Roach J., Oh T., Ho I. Y., Wong M., Detter C., Verhoef F., Predki P., Tay A., Lucas S., Richardson P., Smith S. F., Clark M. S., Edwards Y. J., Doggett N., Zharkikh A., Tavtigian S. V., Pruss D., Barnstead M., Evans C., Baden H., Powell J., Glusman G., Rowen L., Hood L., Tan Y. H., Elgar G., Hawkins T., Venkatesh B., Rokhsar D., Brenner S., Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes, *Science*, 2002, 297, 1301-1310
[21] Kasahara M., Naruse K., Sasaki S., Nakatani Y., Qu W., Ahsan B., Yamada T., Nagayasu Y., Doi K., Kasai Y., Jindo T., Kobayashi D., Shimada A., Toyoda A., Kuroki Y., Fujiyama A., Sasaki T., Shimizu A., Asakawa S., Shimizu N., Hashimoto S., Yang J., Lee Y., Matsushima K., Sugano S., Sakaizumi M., Narita T., Ohishi K., Haga S., Ohta F., Nomoto H., Nogata K., Morishita T., Endo T., Shin I. T., Takeda H., Morishita S., Kohara Y., The medaka draft genome and insights into vertebrate genome evolution, *Nature*, 2007, 447, 714-719
[22] Hellsten U., Harland R. M., Gilchrist M. J., Hendrix D., Jurka J., Kapitonov V., Ovcharenko I., Putnam N. H., Shu S., Taher L., Blitz I. L., Blumberg B., Dichmann D. S., Dubchak I., Amaya E., Detter J. C., Fletcher R., Gerhard D. S., Goodstein D., Graves T., Grigoriev I. V., Grimwood J., Kawashima T., Lindquist E., Lucas S. M., Mead P. E., Mitros T., Ogino H., Ohta Y., Poliakov A. V., Pollet N., Robert J., Salamov A., Sater A. K., Schmutz J., Terry A., Vize P. D., Warren W. C., Wells D., Wills A., Wilson R. K., Zimmerman L. B., Zorn A. M., Grainger R., Grammer T., Khokha M. K., Richardson P. M., Rokhsar D. S., The genome of the Western clawed frog Xenopus tropicalis, *Science*, 2010, 328, 633-636
[23] Hillier L. W., Miller W., Birney E., Warren W., Hardison R. C., Ponting C. P., Bork P., Burt D. W., Groenen M. A., Delany M. E., Dodgson J. B., Chinwalla A. T., Cliften P. F., Clifton S. W., Delehaunty K. D., Fronick C., Fulton R. S., Graves T. A., Kremitzki C., Layman D., Magrini V., McPherson J. D., Miner T. L., Minx P., Nash W. E., Nhan M. N.,

Nelson J. O., Oddy L. G., Pohl C. S., Randall-Maher J., Smith S. M., Wallis J. W., Yang S. P., Romanov M. N., Rondelli C. M., Paton B., Smith J., Morrice D., Daniels L., Tempest H. G., Robertson L., Masabanda J. S., Griffin D. K., Vignal A., Fillon V., Jacobbson L., Kerje S., Andersson L., Crooijmans R. P., Aerts J., van der Poel J. J., Ellegren H., Caldwell R. B., Hubbard S. J., Grafham D. V., Kierzek A. M., McLaren S. R., Overton I. M., Arakawa H., Beattie K. J., Bezzubov Y., Boardman P. E., Bonfield J. K., Croning M. D., Davies R. M., Francis M. D., Humphray S. J., Scott C. E., Taylor R. G., Tickle C., Brown W. R., Rogers J., Buerstedde J. M., Wilson S. A., Stubbs L., Ovcharenko I., Gordon L., Lucas S., Miller M. M., Inoko H., Shiina T., Kaufman J., Salomonsen J., Skjoedt K., Wong G. K., Wang J., Liu B., Wang J., Yu J., Yang H., Nefedov M., Koriabine M., Dejong P. J., Goodstadt L., Webber C., Dickens N. J., Letunic I., Suyama M., Torrents D., von Mering C., Zdobnov E. M., Makova K., Nekrutenko A., Elnitski L., Eswara P., King D. C., Yang S., Tyekucheva S., Radakrishnan A., Harris R. S., Chiaromonte F., Taylor J., He J., Rijnkels M., Griffiths-Jones S., Ureta-Vidal A., Hoffman M. M., Severin J., Searle S. M., Law A. S., Speed D., Waddington D., Cheng Z., Tuzun E., Eichler E., Bao Z., Flicek P., Shteynberg D. D., Brent M. R., Bye J. M., Huckle E. J., Chatterji S., Dewey C., Pachter L., Kouranov A., Mourelatos Z., Hatzigeorgiou A. G., Paterson A. H., Ivarie R., Brandstrom M., Axelsson E., Backstrom N., Berlin S., Webster M. T., Pourquie O., Reymond A., Ucla C., Antonarakis S. E., Long M., Emerson J. J., Betran E., Dupanloup I., Kaessmann H., Hinrichs A. S., Bejerano G., Furey T. S., Harte R. A., Raney B., Siepel A., Kent W. J., Haussler D., Eyras E., Castelo R., Abril J. F., Castellano S., Camara F., Parra G., Guigo R., Bourque G., Tesler G., Pevzner P. A., Smit A., Fulton L. A., Mardis E. R., Wilson R. K., Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution, *Nature*, 2004, 432, 695-716

[24] Warren W. C., Clayton D. F., Ellegren H., Arnold A. P., Hillier L. W., Kunstner A., Searle S., White S., Vilella A. J., Fairley S., Heger A., Kong L., Ponting C. P., Jarvis E. D., Mello C. V., Minx P., Lovell P., Velho T. A., Ferris M., Balakrishnan C. N., Sinha S., Blatti C., London S. E., Li Y., Lin Y. C., George J., Sweedler J., Southey B., Gunaratne P., Watson M., Nam K., Backstrom N., Smeds L., Nabholz B., Itoh Y., Whitney O., Pfenning A. R., Howard J., Volker M., Skinner B. M., Griffin D. K., Ye L., McLaren W. M., Flicek P., Quesada V., Velasco G., Lopez-Otin C., Puente X. S., Olender T., Lancet D., Smit A. F., Hubley R., Konkel M. K., Walker J. A., Batzer M. A., Gu W., Pollock D. D., Chen L., Cheng Z., Eichler E. E., Stapley J., Slate J., Ekblom R., Birkhead T., Burke T., Burt D., Scharff C., Adam I., Richard H., Sultan M., Soldatov A., Lehrach H., Edwards S. V., Yang S. P., Li X., Graves T., Fulton L., Nelson J., Chinwalla A., Hou S., Mardis E. R., Wilson R. K., The genome of a songbird, *Nature*, 2010, 464, 757-762

[25] Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., Gocayne J. D., Amanatides P., Ballew R. M., Huson D. H., Wortman J. R., Zhang Q., Kodira C. D., Zheng X. H., Chen L., Skupski M., Subramanian G., Thomas P. D., Zhang J., Gabor Miklos G. L., Nelson C., Broder S., Clark A. G., Nadeau J., McKusick V. A., Zinder N., Levine A. J., Roberts R. J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Di Francesco V., Dunn P., Eilbeck K., Evangelista C., Gabrielian A. E., Gan W., Ge W., Gong F., Gu Z., Guan P.,

Heiman T. J., Higgins M. E., Ji R. R., Ke Z., Ketchum K. A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G. V., Milshina N., Moore H. M., Naik A. K., Narayan V. A., Neelam B., Nusskern D., Rusch D. B., Salzberg S., Shao W., Shue B., Sun J., Wang Z., Wang A., Wang X., Wang J., Wei M., Wides R., Xiao C., Yan C., Yao A., Ye J., Zhan M., Zhang W., Zhang H., Zhao Q., Zheng L., Zhong F., Zhong W., Zhu S., Zhao S., Gilbert D., Baumhueter S., Spier G., Carter C., Cravchik A., Woodage T., Ali F., An H., Awe A., Baldwin D., Baden H., Barnstead M., Barrow I., Beeson K., Busam D., Carver A., Center A., Cheng M. L., Curry L., Danaher S., Davenport L., Desilets R., Dietz S., Dodson K., Doup L., Ferriera S., Garg N., Gluecksmann A., Hart B., Haynes J., Haynes C., Heiner C., Hladun S., Hostin D., Houck J., Howland T., Ibegwam C., Johnson J., Kalush F., Kline L., Koduru S., Love A., Mann F., May D., McCawley S., McIntosh T., McMullen I., Moy M., Moy L., Murphy B., Nelson K., Pfannkoch C., Pratts E., Puri V., Qureshi H., Reardon M., Rodriguez R., Rogers Y. H., Romblad D., Ruhfel B., Scott R., Sitter C., Smallwood M., Stewart E., Strong R., Suh E., Thomas R., Tint N. N., Tse S., Vech C., Wang G., Wetter J., Williams S., Williams M., Windsor S., Winn-Deen E., Wolfe K., Zaveri J., Zaveri K., Abril J. F., Guigo R., Campbell M. J., Sjolander K. V., Karlak B., Kejariwal A., Mi H., Lazareva B., Hatton T., Narechania A., Diemer K., Muruganujan A., Guo N., Sato S., Bafna V., Istrail S., Lippert R., Schwartz R., Walenz B., Yooseph S., Allen D., Basu A., Baxendale J., Blick L., Caminha M., Carnes-Stine J., Caulk P., Chiang Y. H., Coyne M., Dahlke C., Mays A., Dombroski M., Donnelly M., Ely D., Esparham S., Fosler C., Gire H., Glanowski S., Glasser K., Glodek A., Gorokhov M., Graham K., Gropman B., Harris M., Heil J., Henderson S., Hoover J., Jennings D., Jordan C., Jordan J., Kasha J., Kagan L., Kraft C., Levitsky A., Lewis M., Liu X., Lopez J., Ma D., Majoros W., McDaniel J., Murphy S., Newman M., Nguyen T., Nguyen N., Nodell M., Pan S., Peck J., Peterson M., Rowe W., Sanders R., Scott J., Simpson M., Smith T., Sprague A., Stockwell T., Turner R., Venter E., Wang M., Wen M., Wu D., Wu M., Xia A., Zandieh A., Zhu X., The sequence of the human genome, *Science*, 2001, 291, 1304-1351

[26] Waterston R. H., Lindblad-Toh K., Birney E., Rogers J., Abril J. F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., Antonarakis S. E., Attwood J., Baertsch R., Bailey J., Barlow K., Beck S., Berry E., Birren B., Bloom T., Bork P., Botcherby M., Bray N., Brent M. R., Brown D. G., Brown S. D., Bult C., Burton J., Butler J., Campbell R. D., Carninci P., Cawley S., Chiaromonte F., Chinwalla A. T., Church D. M., Clamp M., Clee C., Collins F. S., Cook L. L., Copley R. R., Coulson A., Couronne O., Cuff J., Curwen V., Cutts T., Daly M., David R., Davies J., Delehaunty K. D., Deri J., Dermitzakis E. T., Dewey C., Dickens N. J., Diekhans M., Dodge S., Dubchak I., Dunn D. M., Eddy S. R., Elnitski L., Emes R. D., Eswara P., Eyras E., Felsenfeld A., Fewell G. A., Flicek P., Foley K., Frankel W. N., Fulton L. A., Fulton R. S., Furey T. S., Gage D., Gibbs R. A., Glusman G., Gnerre S., Goldman N., Goodstadt L., Grafham D., Graves T. A., Green E. D., Gregory S., Guigo R., Guyer M., Hardison R. C., Haussler D., Hayashizaki Y., Hillier L. W., Hinrichs A., Hlavina W., Holzer T., Hsu F., Hua A., Hubbard T., Hunt A., Jackson I., Jaffe D. B., Johnson L. S., Jones M., Jones T. A., Joy A., Kamal M., Karlsson E. K., Karolchik D., Kasprzyk A., Kawai J., Keibler E., Kells C., Kent W. J., Kirby A., Kolbe D. L., Korf I., Kucherlapati R. S., Kulbokas E. J., Kulp D., Landers T., Leger J. P., Leonard S., Letunic I., Levine R., Li J., Li M., Lloyd C., Lucas S., Ma B., Maglott D. R., Mardis E. R., Matthews L., Mauceli E., Mayer J. H., McCarthy M., McCombie W. R., McLaren S., McLay K., McPherson J. D., Meldrim J., Meredith B., Mesirov J. P., Miller W., Miner T. L.,

Mongin E., Montgomery K. T., Morgan M., Mott R., Mullikin J. C., Muzny D. M., Nash W. E., Nelson J. O., Nhan M. N., Nicol R., Ning Z., Nusbaum C., O'Connor M. J., Okazaki Y., Oliver K., Overton-Larty E., Pachter L., Parra G., Pepin K. H., Peterson J., Pevzner P., Plumb R., Pohl C. S., Poliakov A., Ponce T. C., Ponting C. P., Potter S., Quail M., Reymond A., Roe B. A., Roskin K. M., Rubin E. M., Rust A. G., Santos R., Sapojnikov V., Schultz B., Schultz J., Schwartz M. S., Schwartz S., Scott C., Seaman S., Searle S., Sharpe T., Sheridan A., Shownkeen R., Sims S., Singer J. B., Slater G., Smit A., Smith D. R., Spencer B., Stabenau A., Stange-Thomann N., Sugnet C., Suyama M., Tesler G., Thompson J., Torrents D., Trevaskis E., Tromp J., Ucla C., Ureta-Vidal A., Vinson J. P., Von Niederhausern A. C., Wade C. M., Wall M., Weber R. J., Weiss R. B., Wendl M. C., West A. P., Wetterstrand K., Wheeler R., Whelan S., Wierzbowski J., Willey D., Williams S., Wilson R. K., Winter E., Worley K. C., Wyman D., Yang S., Yang S. P., Zdobnov E. M., Zody M. C., Lander E. S., Initial sequencing and comparative analysis of the mouse genome, *Nature*, 2002, 420, 520-562

[27] Gibbs R. A., Weinstock G. M., Metzker M. L., Muzny D. M., Sodergren E. J., Scherer S., Scott G., Steffen D., Worley K. C., Burch P. E., Okwuonu G., Hines S., Lewis L., DeRamo C., Delgado O., Dugan-Rocha S., Miner G., Morgan M., Hawes A., Gill R., Celera, Holt R. A., Adams M. D., Amanatides P. G., Baden-Tillson H., Barnstead M., Chin S., Evans C. A., Ferriera S., Fosler C., Glodek A., Gu Z., Jennings D., Kraft C. L., Nguyen T., Pfannkoch C. M., Sitter C., Sutton G. G., Venter J. C., Woodage T., Smith D., Lee H. M., Gustafson E., Cahill P., Kana A., Doucette-Stamm L., Weinstock K., Fechtel K., Weiss R. B., Dunn D. M., Green E. D., Blakesley R. W., Bouffard G. G., De Jong P. J., Osoegawa K., Zhu B., Marra M., Schein J., Bosdet I., Fjell C., Jones S., Krzywinski M., Mathewson C., Siddiqui A., Wye N., McPherson J., Zhao S., Fraser C. M., Shetty J., Shatsman S., Geer K., Chen Y., Abramzon S., Nierman W. C., Havlak P. H., Chen R., Durbin K. J., Egan A., Ren Y., Song X. Z., Li B., Liu Y., Qin X., Cawley S., Worley K. C., Cooney A. J., D'Souza L. M., Martin K., Wu J. Q., Gonzalez-Garay M. L., Jackson A. R., Kalafus K. J., McLeod M. P., Milosavljevic A., Virk D., Volkov A., Wheeler D. A., Zhang Z., Bailey J. A., Eichler E. E., Tuzun E., Birney E., Mongin E., Ureta-Vidal A., Woodwark C., Zdobnov E., Bork P., Suyama M., Torrents D., Alexandersson M., Trask B. J., Young J. M., Huang H., Wang H., Xing H., Daniels S., Gietzen D., Schmidt J., Stevens K., Vitt U., Wingrove J., Camara F., Mar Alba M., Abril J. F., Guigo R., Smit A., Dubchak I., Rubin E. M., Couronne O., Poliakov A., Hubner N., Ganten D., Goesele C., Hummel O., Kreitler T., Lee Y. A., Monti J., Schulz H., Zimdahl H., Himmelbauer H., Lehrach H., Jacob H. J., Bromberg S., Gullings-Handley J., Jensen-Seaman M. I., Kwitek A. E., Lazar J., Pasko D., Tonellato P. J., Twigger S., Ponting C. P., Duarte J. M., Rice S., Goodstadt L., Beatson S. A., Emes R. D., Winter E. E., Webber C., Brandt P., Nyakatura G., Adetobi M., Chiaromonte F., Elnitski L., Eswara P., Hardison R. C., Hou M., Kolbe D., Makova K., Miller W., Nekrutenko A., Riemer C., Schwartz S., Taylor J., Yang S., Zhang Y., Lindpaintner K., Andrews T. D., Caccamo M., Clamp M., Clarke L., Curwen V., Durbin R., Eyras E., Searle S. M., Cooper G. M., Batzoglou S., Brudno M., Sidow A., Stone E. A., Venter J. C., Payseur B. A., Bourque G., Lopez-Otin C., Puente X. S., Chakrabarti K., Chatterji S., Dewey C., Pachter L., Bray N., Yap V. B., Caspi A., Tesler G., Pevzner P. A., Haussler D., Roskin K. M., Baertsch R., Clawson H., Furey T. S., Hinrichs A. S., Karolchik D., Kent W. J., Rosenbloom K. R., Trumbower H., Weirauch M., Cooper D. N., Stenson P. D., Ma B., Brent M., Arumugam M., Shteynberg D., Copley R. R., Taylor M. S.,

Riethman H., Mudunuri U., Peterson J., Guyer M., Felsenfeld A., Old S., Mockrin S., Collins F., Genome sequence of the Brown Norway rat yields insights into mammalian evolution, *Nature*, 2004, 428, 493-521
[28] Putnam N. H., Butts T., Ferrier D. E., Furlong R. F., Hellsten U., Kawashima T., Robinson-Rechavi M., Shoguchi E., Terry A., Yu J. K., Benito-Gutierrez E. L., Dubchak I., Garcia-Fernandez J., Gibson-Brown J. J., Grigoriev I. V., Horton A. C., de Jong P. J., Jurka J., Kapitonov V. V., Kohara Y., Kuroki Y., Lindquist E., Lucas S., Osoegawa K., Pennacchio L. A., Salamov A. A., Satou Y., Sauka-Spengler T., Schmutz J., Shin I. T., Toyoda A., Bronner-Fraser M., Fujiyama A., Holland L. Z., Holland P. W., Satoh N., Rokhsar D. S., The amphioxus genome and the evolution of the chordate karyotype, *Nature*, 2008, 453, 1064-1071
[29] Putnam N. H., Srivastava M., Hellsten U., Dirks B., Chapman J., Salamov A., Terry A., Shapiro H., Lindquist E., Kapitonov V. V., Jurka J., Genikhovich G., Grigoriev I. V., Lucas S. M., Steele R. E., Finnerty J. R., Technau U., Martindale M. Q., Rokhsar D. S., Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization, *Science*, 2007, 317, 86-94
[30] Adams M. D., Celniker S. E., Holt R. A., Evans C. A., Gocayne J. D., Amanatides P. G., Scherer S. E., Li P. W., Hoskins R. A., Galle R. F., George R. A., Lewis S. E., Richards S., Ashburner M., Henderson S. N., Sutton G. G., Wortman J. R., Yandell M. D., Zhang Q., Chen L. X., Brandon R. C., Rogers Y. H., Blazej R. G., Champe M., Pfeiffer B. D., Wan K. H., Doyle C., Baxter E. G., Helt G., Nelson C. R., Gabor G. L., Abril J. F., Agbayani A., An H. J., Andrews-Pfannkoch C., Baldwin D., Ballew R. M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E. M., Beeson K. Y., Benos P. V., Berman B. P., Bhandari D., Bolshakov S., Borkova D., Botchan M. R., Bouck J., Brokstein P., Brottier P., Burtis K. C., Busam D. A., Butler H., Cadieu E., Center A., Chandra I., Cherry J. M., Cawley S., Dahlke C., Davenport L. B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A. D., Dew I., Dietz S. M., Dodson K., Doup L. E., Downes M., Dugan-Rocha S., Dunkov B. C., Dunn P., Durbin K. J., Evangelista C. C., Ferraz C., Ferriera S., Fleischmann W., Fosler C., Gabrielian A. E., Garg N. S., Gelbart W. M., Glasser K., Glodek A., Gong F., Gorrell J. H., Gu Z., Guan P., Harris M., Harris N. L., Harvey D., Heiman T. J., Hernandez J. R., Houck J., Hostin D., Houston K. A., Howland T. J., Wei M. H., Ibegwam C., Jalali M., Kalush F., Karpen G. H., Ke Z., Kennison J. A., Ketchum K. A., Kimmel B. E., Kodira C. D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A. A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T. C., McLeod M. P., McPherson D., Merkulov G., Milshina N. V., Mobarry C., Morris J., Moshrefi A., Mount S. M., Moy M., Murphy B., Murphy L., Muzny D. M., Nelson D. L., Nelson D. R., Nelson K. A., Nixon K., Nusskern D. R., Pacleb J. M., Palazzolo M., Pittman G. S., Pan S., Pollard J., Puri V., Reese M. G., Reinert K., Remington K., Saunders R. D., Scheeler F., Shen H., Shue B. C., Siden-Kiamos I., Simpson M., Skupski M. P., Smith T., Spier E., Spradling A. C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A. H., Wang X., Wang Z. Y., Wassarman D. A., Weinstock G. M., Weissenbach J., Williams S. M., WoodageT, Worley K. C., Wu D., Yang S., Yao Q. A., Ye J., Yeh R. F., Zaveri J. S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X. H., Zhong F. N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H. O., Gibbs R. A., Myers E. W., Rubin G. M., Venter J. C., The genome sequence of Drosophila melanogaster, *Science*, 2000, 287, 2185-2195

[31] Genome sequence of the nematode C. elegans: a platform for investigating biology, *Science*, 1998, 282, 2012-2018
[32] Mulder N. J., Apweiler R., Tools and resources for identifying protein families, domains and motifs, *Genome Biol*, 2002, 3, REVIEWS2001
[33] Wernersson R., Pedersen A. G., RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences, *Nucleic Acids Res*, 2003, 31, 3537-3539
[34] Altschul S. F., Gish W., Local alignment statistics, *Methods Enzymol*, 1996, 266, 460-480
[35] Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J., Basic local alignment search tool, *J Mol Biol*, 1990, 215, 403-410
[36] Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res*, 1997, 25, 3389-3402
[37] Pearson W. R., Rapid and sensitive sequence comparison with FASTP and FASTA, *Methods Enzymol*, 1990, 183, 63-98
[38] Pearson W. R., Lipman D. J., Improved tools for biological sequence comparison, *Proc Natl Acad Sci U S A*, 1988, 85, 2444-2448
[39] McGuffin L. J., Bryson K., Jones D. T., The PSIPRED protein structure prediction server, *Bioinformatics*, 2000, 16, 404-405
[40] Ward J. J., McGuffin L. J., Bryson K., Buxton B. F., Jones D. T., The DISOPRED server for the prediction of protein disorder, *Bioinformatics*, 2004, 20, 2138-2139
[41] Saitou N., Nei M., The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol Biol Evol*, 1987, 4, 406-425
[42] Felsenstein J., Confidence Limits on Phylogenies: An Approach Using the Bootstrap., *Evolution*, 1985, 39, 783-791
[43] Tamura K., Dudley J., Nei M., Kumar S., MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol Biol Evol*, 2007, 24, 1596-1599
[44] Flicek P., Aken B. L., Ballester B., Beal K., Bragin E., Brent S., Chen Y., Clapham P., Coates G., Fairley S., Fitzgerald S., Fernandez-Banet J., Gordon L., Graf S., Haider S., Hammond M., Howe K., Jenkinson A., Johnson N., Kahari A., Keefe D., Keenan S., Kinsella R., Kokocinski F., Koscielny G., Kulesha E., Lawson D., Longden I., Massingham T., McLaren W., Megy K., Overduin B., Pritchard B., Rios D., Ruffier M., Schuster M., Slater G., Smedley D., Spudich G., Tang Y. A., Trevanion S., Vilella A., Vogel J., White S., Wilder S. P., Zadissa A., Birney E., Cunningham F., Dunham I., Durbin R., Fernandez-Suarez X. M., Herrero J., Hubbard T. J., Parker A., Proctor G., Smith J., Searle S. M., Ensembl's 10th year, *Nucleic Acids Res*, 2010, 38, D557-562
[45] Hubbard T. J., Aken B. L., Ayling S., Ballester B., Beal K., Bragin E., Brent S., Chen Y., Clapham P., Clarke L., Coates G., Fairley S., Fitzgerald S., Fernandez-Banet J., Gordon L., Graf S., Haider S., Hammond M., Holland R., Howe K., Jenkinson A., Johnson N., Kahari A., Keefe D., Keenan S., Kinsella R., Kokocinski F., Kulesha E., Lawson D., Longden I., Megy K., Meidl P., Overduin B., Parker A., Pritchard B., Rios D., Schuster M., Slater G., Smedley D., Spooner W., Spudich G., Trevanion S., Vilella A., Vogel J., White S., Wilder S., Zadissa A., Birney E., Cunningham F., Curwen V., Durbin R., Fernandez-Suarez X. M., Herrero J., Kasprzyk A., Proctor G., Smith J., Searle S., Flicek P., Ensembl 2009, *Nucleic Acids Res*, 2009, 37, D690-697

[46] Dunker A. K., Brown C. J., Lawson J. D., Iakoucheva L. M., Obradovic Z., Intrinsic disorder and protein function, *Biochemistry*, 2002, 41, 6573-6582
[47] Dunker A. K., Brown C. J., Obradovic Z., Identification and functions of usefully disordered proteins, *Adv Protein Chem*, 2002, 62, 25-49
[48] Dyson H. J., Wright P. E., Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol*, 2005, 6, 197-208
[49] Uversky V. N., Natively unfolded proteins: a point where biology waits for physics, *Protein Sci*, 2002, 11, 739-756
[50] Iakoucheva L. M., Brown C. J., Lawson J. D., Obradovic Z., Dunker A. K., Intrinsic disorder in cell-signaling and cancer-associated proteins, *J Mol Biol*, 2002, 323, 573-584
[51] Dunker A. K., Cortese M. S., Romero P., Iakoucheva L. M., Uversky V. N., Flexible nets. The roles of intrinsic disorder in protein interaction networks, *FEBS J*, 2005, 272, 5129-5148
[52] Uversky V. N., Oldfield C. J., Dunker A. K., Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, *J Mol Recognit*, 2005, 18, 343-384
[53] Garza A. S., Ahmad N., Kumar R., Role of intrinsically disordered protein regions/domains in transcriptional regulation, *Life Sci*, 2009, 84, 189-193
[54] Uversky V. N., Gillespie J. R., Fink A. L., Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins*, 2000, 41, 415-427
[55] Rogers A., Antoshechkin I., Bieri T., Blasiar D., Bastiani C., Canaran P., Chan J., Chen W. J., Davis P., Fernandes J., Fiedler T. J., Han M., Harris T. W., Kishore R., Lee R., McKay S., Muller H. M., Nakamura C., Ozersky P., Petcherski A., Schindelman G., Schwarz E. M., Spooner W., Tuli M. A., Van Auken K., Wang D., Wang X., Williams G., Yook K., Durbin R., Stein L. D., Spieth J., Sternberg P. W., WormBase 2007, *Nucleic acids research*, 2008, 36, D612-617
[56] Wheeler D. L., Barrett T., Benson D. A., Bryant S. H., Canese K., Chetvernin V., Church D. M., DiCuccio M., Edgar R., Federhen S., Geer L. Y., Helmberg W., Kapustin Y., Kenton D. L., Khovayko O., Lipman D. J., Madden T. L., Maglott D. R., Ostell J., Pruitt K. D., Schuler G. D., Schriml L. M., Sequeira E., Sherry S. T., Sirotkin K., Souvorov A., Starchenko G., Suzek T. O., Tatusov R., Tatusova T. A., Wagner L., Yaschenko E., Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res*, 2006, 34, D173-180
[57] Nuzhdin S., Wayne M., Harmon K., McIntyre L., Common Pattern of Evolution of Gene Expression Level and Protein Sequence in Drosophila, *Molecular Biology and Evolution*, 2004, 21, 1308-1317
[58] Barbour K., Goodwin R., Guillonneau F., Wang Y., Baumann H., Berger F., Functional Diversification During Evolution of the Murine {alpha}1-Proteinase Inhibitor Family: Role of the Hypervariable Reactive Center Loop, *Molecular Biology and Evolution*, 2002, 19, 718-727
[59] Buck M., Atchley W., Networks of Coevolving Sites in Structural and Functional Domains of Serpin Proteins, *Molecular Biology and Evolution*, 2005, 22, 1627-1634
[60] Wolfsberg T. G., Using the NCBI map viewer to browse genomic sequence data, *Curr Protoc Bioinformatics*, 2010, Chapter 1, Unit 1 5 1-25

[61] Karolchik D., Bejerano G., Hinrichs A. S., Kuhn R. M., Miller W., Rosenbloom K. R., Zweig A. S., Haussler D., Kent W. J., Comparative genomic analysis using the UCSC genome browser, *Methods in molecular biology (Clifton, NJ*, 2007, 395, 17-34
[62] Higgins D. G., Thompson J. D., Gibson T. J., Using CLUSTAL for multiple sequence alignments, *Methods Enzymol*, 1996, 266, 383-402
[63] Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G., The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res*, 1997, 25, 4876-4882
[64] Edgar R. C., MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinformatics*, 2004, 5, 113
[65] Edgar R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res*, 2004, 32, 1792-1797
[66] Nicholas K. B., Nicholas H.B. Jr., and Deerfield D. W. I., GeneDoc: Analysis and Visualization of Genetic Variation, *EMBNEWNEWS*, 1997, 4, 14
[67] Waterhouse A. M., Procter J. B., Martin D. M., Clamp M., Barton G. J., Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, 2009, 25, 1189-1191
[68] Clamp M., Cuff J., Searle S. M., Barton G. J., The Jalview Java alignment editor, *Bioinformatics*, 2004, 20, 426-427
[69] Crooks G. E., Hon G., Chandonia J. M., Brenner S. E., WebLogo: a sequence logo generator, *Genome Res*, 2004, 14, 1188-1190
[70] Blair Hedges S., Kumar S., Genomic clocks and evolutionary timescales, *Trends Genet*, 2003, 19, 200 - 206
[71] Ponting C. P., The functional repertoires of metazoan genomes, *Nat Rev Genet*, 2008, 9, 689-698

## Note

This article was originally presented at the National Symposium on Modern Approaches and Innovations in Biotechnology, Nov 2010, Meerut, India.